PCT/DK97/00342

09/242657

Kongeriget Danmark

Patent application No.: 0886/96

Date of filing: 23 Aug 1996

Applicant: Peter Ruhdal Jensen, Søgårdsvej 19, DK-2820 Gentofte, DK

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims and drawing as filed with the application on the filing date indicated above.
- Letter and correction sheets of page 27 and 34 as filed on 22 Oct 1996.

Patentdirektoratet · Kongeriget Danmark ·

Erhvervsministeriet
Patentdirektoratet

TAASTRUP 02 Sep 1997

Jytte Hansen
Kontorfuldmægtig

0886/9623 AUG. 96

Hofman-Bang & Boutard, Lehmann & Ree A/S
Patents • Trademarks • Designs

Peter Ruhdal Jensen
Søgårdsvej 19
DK-2820 Gentofte

Deres ref.:	Vor ref.:	Dato:
-	DK 2960322 JNY	23. august 1996

Kunstige promotorbiblioteker for mikroorganismer

☑ Adelgade 15
DK-1304 København K
Danmark
Telefon: +45 33 15 05 85
Telefax: +45 33 15 75 85
Telex: 19 085 hbb dk

☐ Grundtvigsvej 37
DK-1864 Frederiksberg C
Danmark
Telefon: +45 31 31 32 31
Telefax: +45 31 31 12 90
Telex: 15 067 lehree dk

☐ Ryesgade 3
Postboks 367
DK-8100 Aarhus C
Danmark
Telefon: +45 86 20 22 22
Telefax: +45 86 20 22 10

Telegram: Hofmanbang
Reg. nr.: 207 235
Giro: 7 02 57 85

Artificial promoter libraries for microorganisms

This invention concerns artificial promoter libraries and a method of constructing an artificial promoter library for a selected microorganism or group of microorganisms. The invention also concerns a method of optimizing the expression of a recombinant heterologous gene in a selected microorganism by use of such an artificial promoter library for that microorganism. In connection with this invention the term "microorganism" shall be taken broadly to include prokaryotic organisms such as bacteria as well as eukaryotic microorganisms such as yeasts, other fungi and cell lines of higher organisms.

BACKGROUND OF THE INVENTION

Metabolic engineering of microorganisms is still in its "infancy" with respect to industrial applications, despite the fact that genetic engineering has now been feasible for more than a decade. To a large extent, this may be due to the disappointing outcome of many of the attempts so far to improve strain performance. There are at least two reasons for the negative outcome of the attempts to increase metabolic fluxes:

One is that the genetic engineer tends to overlook the subtlety of control and regulation of cellular metabolism. The expression of enzymes that are expected to be "rate limiting" are increased 10 to 100 fold, e.g. by placing the gene on a high copy number plasmid. Or, a branching flux in a pathway is eliminated by deleting a gene. Quite often, this will have secondary effects on the metabolism, for instance by lowering metabolite concentrations that are essential to other parts of the cellular metabolism (e.g. processes that are essential to the growth of the organsim) and the net result may be that the overall performance of the cell with respect to

the desired product is decreased. Instead, it is necessary to "tune" the expression of the relevant gene around the normal expression level and determine the optimal expression level, as the level that maximizes the flux.

The second reason for the negative outcome lies in the "rate limiting" concept itself: both metabolic control theory (Kacser and Burns, 1973) and experimental determinations of control by individual steps in a pathway (Schaaff *et al.*, 1989; Jensen *et al.*, 1993) have shown that reaction steps which were expected to be "rate limiting" with respect to a particular flux, turned out to have no or very little control over the flux. Instead, the control and regulation of the cellular metabolism turned out to be distributed over several enzymes in a pathway, and it may be necessary to enhance the expression of several enzymes in order to obtain a higher flux.

According to metabolic control theory, the total flux control exerted by all the enzymes in a pathway, should always sum up to 1. Therefore, after one enzyme concentration has been optimized, the flux control will have shifted to another enzyme(s), and it may then be useful to perform additional rounds of enzyme optimization in order to increase the flux further.

In summary, flux optimization requires 1) fine-tuning of enzyme concentration rather than many fold overexpression and often 2) optimization of the level of several enzymes in a pathway rather than looking for the "rate limiting" step.

There are now many systems available that allow one to increase the gene expression more than 1000 fold and/or to turn on gene expression at a particular time point during a fermentation process (e.g. using temperature inducible systems). With respect to tuning gene expression

in the fermenter, to say 150% or 70% of the normal expression level, it becomes more difficult. In principle, one could use a *lac*-type promoter in front of the gene of interest, and then add a certain amount of an inducer of the *lac* system, for instance IPTG (isopropyl-ß-D-thiogalactoside), or use a temperature sensitive system at the correct temperature. These possibilities are often not practical for large scale industrial applications. The alternative is to use a promoter that has exactly the right strength. However, such promoters are seldom available, and furthermore one needs a range of promoter activities in order to optimize the expression of the gene in the first place, see below.

SUMMARY OF THE INVENTION

The present invention provides an artificial promoter library for a selected microorganism or group of microorganisms, comprising a mixture of double stranded DNA fragments the sense strands of which comprise at least one consensus sequence or part(s) thereof of efficient promoters from said microorganism or group of microorganisms and surrounding or intermediate nucleotide sequences (spacers) of variable length in which the nucleotides are selected randomly among the nucleobases A, T, C and G.

The sense strands of the double stranded DNA fragments may also include a regulatory DNA sequence imparting a specific regulatory feature to the promoters of the library. Such specific regulatory feature is preferably activation by a change in the growth conditions, such as a change in the pH, osmolarity, temperature or growth phase.

For cloning purposes the double stranded DNA fragments usually have sequences comprising one or more recognition sites for restriction endonucleases added to their ends;

most conveniently sequences specifying multiple recognition sites for restriction endonucleases (multiple cloning sites MCS).

The selected microrganism or group of microorganisms may be selected from prokaryotes and from eukaryotic microorganisms such as yeasts, other fungi and cell lines of higher organisms.

An interesting group of prokaryotes i.a. in the dairy industry consist of lactic acid bacteria of the genus *Lactococcus* (lactococci), in particular strains of the species *Lactococcus lactis*.

In an artificial promoter library for lactococci said consensus sequences should comprise the -35 signal (-35 to -30): TTGACA or the -10 signal (-12 to -7): TATAAT or a part of one of them comprising at least 4 conserved nucleotides.

More efficient promoters are usually obtained when said consensus sequences comprise the -35 signal (-35 to -30): TTGACA and the -10 signal (-12 to -7): TATAAT or parts of both comprising together at least 6 conserved nucleotides; and the most efficient promoters are obtained when said consensus sequences further comprise intervening conserved motifs, e.g. selected from the conserved motifs -44 to -41: AGTT, -40 to -36: TATTC and +1 to +8: GTACTGTT.

In such promoters the length of the spacer between the -35 signal and the -10 signal shoyuld be 14-23 bp, preferably 16-18 bp, and more preferably 17 bp.

An interesting eukaryotic microorganism is the yeast species *Saccharomyces cerevisiae*, normal baker's yeast.

In *Saccharomyces cerevisiae* said consensus sequences should normally comprise a transcription initiation signal (TI box) functioning in *Saccharomyces cerevisiae*. In most cases they should also comprise a TATA box and/or at least one upstream activation sequence (UAS) upstream of the TI box.

In a specific embodiment of an artificial promoter library according to the invention for *Saccharomyces cerevisiae* said consensus sequences comprise the TI box: CTCTTAAGTGCAAGTGACTGCGA, which also functions as the binding site for the arginine repressor, *argR*, the TATA box: TATAAA, and the UAS_{GCN4p}: TGACTCA.

The present invention also provides a method of constructing an artificial promoter library for a selected microorganism or group of microorganisms, which comprises selecting at least one consensus sequence of efficient promoters from said microorganism or group of microorganisms; synthesizing a mixture of single stranded DNA sequences comprising said consensus sequence(s) or part(s) thereof and surrounding or intermediate nucleotide sequences (spacers)of variable length in which the nucleotides are selected randomly among the nucleobases A, T, C and G; and converting the single stranded DNA sequences into double stranded DNA fragments by second strand synthesis.

In order to obtain an artificial promoter library which is susceptible to regulation, the single stranded DNA sequences which are synthesized may include a regulatory DNA sequence imparting a specific regulatory feature to the promoters of the library. Such specific regulatory feature is preferably activation by a change in the growth conditions, such as a change in the pH, osmolarity, temperature or growth phase.

Also, in order to obtain an artificial promoter library suitable for cloning, sequences specifying one or more recognition sites for restriction endonucleases may be added to the ends of the single stranded DNA sequences in the synthesis, or linkers comprising such restriction sites may be ligated to the ends of the double stranded DNA fragments. Most conveniently such sequences specify multiple recognition sites for restriction endonucleases (multiple cloning sites MCS).

The selected microorganisms for which artificial promoter libraries can be prepared by the method according to the invention and the various degenerated sequences to be chosen for the promoter libraries of specific microorganisms are the same as discussed above for the artificial promoter libraries per se.

With respect to possible uses of the artificial promoter libraries described above, the invention further provides a method of optimizing the expression of a recombinant heterologous gene in a selected microorganism, which comprises transforming or transfecting the microorganism with a set of vectors each including said heterologous gene under the control of at least one member of an artificial promoter library according to any one of claims 1-24 or constructed by the method according to any one of claims 25-48, said set of vectors covering a wide range of promoter activities in relatively small steps, growing the selected clones and screening them to find the one showing maximum metabolic flux of the product encoded by said heterologous gene.

DETAILED DESCRIPTION OF THE INVENTION

The features that make a promoter function efficiently in a particular microorganism are the consensus sequences (e.g. the -10 region, -35 region, etc.) and the optimal

distances between these. According to the literature, by including these elements, the resulting promoters would tend to become strong, and the sequence of the spacers between the consensus sequences should be less important for the strength of the promoters. In principle, modulation of the strength of these promoters could then be achieved by base pair changes in the consensus sequences. The promoter libraries that we wish to construct should cover a range of promoter activity from, say less than 1 up to several thousand relative units, in small steps, for example an increase in activity by 50-100% per step. But the impact of changes in the consensus sequences on the promoter strength will tend to be large because relatively few base pairs determine the strength, and therefore it will not be feasible to achieve these small steps of strength modulation.

We have used a different approach to obtain sets of promoters, the strength of which covers (in small steps) the range of gene expressions that could become of interest. First, degenerated oligonucleotides, approximately 100 nucleotides long, are synthesized for the microorganism for which one wishes to construct a promoter library. The sequences of the oligonucleotides are written by combining as much of the available knowledge from the literature as possible, on the features that makes a promoter function efficiently in that particular microorganism. Secondly, the single stranded oligonucleotides are converted into double stranded DNA fragments and cloned into a promoter probing vector. The promoter-containing clones are identified for their ability to give colonies with various extents of colour on indicator plates. This should in principle give us only very strong promoters, but we discovered that by allowing the spacer sequences between the consensus sequences to vary in a random manner, the strength of the resulting promoters are modulated. In fact, using this method we obtained promoter

libraries, spanning the entire range of promoter activities that can become of interest, in small steps of activity change.

Optimization of gene expression could then be achieved as follows: 1) From the promoter library one chooses promoters that have e.g. 25%, 50%, 100%, 200% and 400% of the strength of the wild-type promoter. 2) Then, these promoters are cloned in place of the wild-type promoter upstream of the gene of interest. 3) The magnitude of the variable to be optimized (e.g. the flux through a pathway) that is obtained with each of these five constructs is then measured and the optimal construct is used directly as the production organism. It may be necessary to fine-tune the expression further or to expand the range of promoter activities. A direct advantage of this system over the inducible systems described above is that once the optimal promoter activity has been determined, the strain is in principle ready for use in the fermentation process.

Often it is desirable to activate gene expression to a certain extent and at a certain stage of a fermentation, e.g. because the gene product that is expressed inhibits the growth of the cells. It is then useful to combine the above technique for obtaining promoters of different strength with some regulatory mechanism, e.g. so that the promoter will be activated by a change in the pH, temperature or growth phase.

Thus, in a specific embodiment of the invention, as illustrated in Example 2, the above approach is used in combination with specific regulatory DNA sequences to generate a library of promoters that have a specific regulatory feature in common.

In addition to prokaryotes, eukaryotes (yeast and other fungi as well as mammalian cell lines,) are important microorganisms for production of a range of organic compounds and various proteins. It is therefore of interest to develop the above approach for modulating gene expression in these organisms as well. Thus, in another specific embodiment of the invention, as illustrated in Example 3, the the above technique for obtaining promoters of different strength is elaborated for the bakers yeast, *Saccharomyces cerevisiae*.

EXAMPLE 1

Design of a degenerated oligonucleotide for a *L. lactis* promoter library.

According to the literature (see review in de Vos & Simons, 1994), strong promoters in *L. lactis* tend to have the following nucleotide sequences in common (numbers refer to the position relative to the transcription initiation site, which is given number +1): -12 to -7: TATAAT; -15 to -14: TG; -35 to -30: TTGACA. The spacing between -10 and -35 seems to be 17 nucleotides. However, closer comparison of the promoter sequences that have been published for *L. lactis* reveals that in a number of positions besides the ones mentioned above, nucleotides are more or less well conserved. Some of these positions are: -1: A; -3: A or T (=W); -6: A; -13: A or G (=R); -40 to -36: TATTC. In addition, Nilsson and Johansen (1994, BBA) pointed out two motives, +1 to +8: GTACTGTT, and -44 to -41: AGTT, that appear to be well conserved between relatively strong promoters (promoters for transfer RNA and ribosomal RNA operons) from *L. lactis*. These motives may confer both strength and growth rate dependent expression from the promoter.

When these additional motives are included, one arrives at the following 53 nucleobase degenerated sequence for an efficient promoter in *L. lactis*. Out of these 53 nucleobases, 34 bases are conserved, two are semi-conserved (R and W) and 17 are allowed to vary randomly between the four nucleobases.

5'AGTTTATTCTTGACANNNNNNNNNNNNNNTGRTATAATANNWNAGTACTGTT 3'

In addition, this degenerated sequence is flanked by sequences that specify multiple recognition sites for restriction endonucleases (multiple cloning site MCS), and the actual oligonucleotide mixture to be synthesized has the following degenerated sequence reported in SEQ ID No. 1:

```
      MboI
      DpnI
      AlwI
     HlaIV
     BstYI
     BamHI MseI
     AlwI AflII SspI  NsiI
  1
5'CGGGATCCTTAAGAATATTATGCATNNNNNNAGTTTATTCTTGACANNNNNNNNNNNNNNT

                                   AluI
                                  PvuII
                                  NspBII
                                SfcI
                            MscI   Fnu4HI
                       RsaI HpaI PstI
                      ScaI  HincII BbvI  EcoRI
61                                          100
 GRTATAATANNWNAGTACTGTTAACTGCAGCTGAATTCGG 3'
```

A mixture of oligonucleotides according to this specification was synthesized by Hobolth DNA synthesis.

This oligonucleotide mixture is single stranded initially and must be converted into double stranded DNA fragments for the purpose of cloning. This was done by synthesizing a 10 base pair oligonucleotide, having a sequence complementary to the 3' end of the promoter oligonucleotide. This oligonucleotide was then used as a primer for the

second strand synthesis by the Klenow fragment of DNA polymerase I in the presence of dATP, dCTP, dGTP and dTTP. In this way the second DNA strand became exactly complementary to the first DNA strand.

The result is a mixture of 100 bp double stranded DNA fragments containing multiple recognition sites for restriction endonucleases in both the 3' and 5' end. These DNA fragments were digested with restriction endonucleases, in order to create DNA ends compatible with the ends of the vector DNA fragment, pAK80 (Israelsen *et al.*, 1995). pAK80 is a "shuttle vector", meaning that it has replication origins for propagation in both *E. coli* and *L. lactis*. In this way, the cloning steps can be conveniently performed in *E. coli*, while the subsequent physiological experiments can be done in *L. lactis*. Furthermore, pAK80 carries a promoterless betagalactosidase reporter gene system (*lacLM*) downstream a multiple cloning site. Thus, pAK80 does not express the *lacLM* genes, unless a promoter is inserted in the multiple cloning site.

Two cloning strategies were used for cloning the mixture of double stranded DNA fragments into the cloning vector pAK80:

1) The mixture was digested with *BamHI* and *PstI* and the vector pAK80 with *PstI* and *BglII* (*BglII* is compatible with *BamHI*).

2) The mixture was digested with *SspI* and *HincII* and the vector pAK80 with *SmaI* (all three enzymes produce blunt end DNA fragments).

In both cases the vector DNA was further treated with Calf Intestine Phosphatase (CIP) to prevent religation of the cloning vector. Subsequently, promoter fragment and

vector were ligated overnight at 16 °C using T4 DNA ligase and standard ligation conditions.

Ligation mixtures were transformed into *E. coli* K-12 D*lac*, with selection for erythromycin resistance. Cells of the *E. coli* K-12 strain, MT102, were made competent using standard treatment with Ca^{++} ions (Maniatis et al., 1982). Ligation mixtures were then transformed into these cells using a standard transformation procedure (Maniatis et al., 1982), and the resulting clones were screened for betagalactosidase activity that will produce blue colonies on plates containing X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside). The transformation mixture was plated on LB plates containing 200 µg/ml erythromycin, 1% glycerol and 100 µg/ml X-gal. 150 erythromycin resistant transformants were obtained, all white initially, but after prolonged incubation (two weeks at 4 °C), 46 of these colonies had become light blue. Thus, using strategy 1) we found 17 blue colonies (CP30 to CP46), and using strategy 2) we found 29 blue colonies (CP1 to CP29).

Plasmid DNA was isolated from each of these clones (CP1 to CP46) and analysed by restriction enzyme mapping. Nearly all plasmids contained promoter fragments inserted into the MCS of pAK80, in the oriemtation that would direct transcription of the otherwise promoterless *lacLM* genes on this vector.

These 46 plasmid DNA preparations were then transformed into *L. lactis* subspecies *lactis* MG1363 with selection for erythromycin resistance. Cells of the *L. lactis* subspecies *lactis* strain, MG1363 (Gasson, 1983) were made competent by growing the cells overnight in SGM17 medium, containing 2% glycine, as described by Holo and Nes (1989). Plasmid DNA from each of the 46 clones described above was then transformed into these cells using the electroporation procedure (Holo and Nes, 1989). After re-

generation, the cells were plated on SR plates containing 2 µg/ml erythromycin. Subsequent screening for blue color on X-gal plates revealed large differences in betagalactosidase activity between the 46 clones; some clones gave dark blue colonies after 24 hours of incubation, others only light blue colonies after more than 1 week of incubation.

The betagalactosidase activities of liquid cultures of the 46 clones in MG1363 were then determined as described by Miller (1972) and modified by Israelsen *et al.* (1995). Cultures of the strain MG1363, each carrying one of the 46 plasmid derivatives of pAK80, were grown in M17 medium supplemented with 1% glucose overnight at 30 °C. 25-100 µl of these cultures were then used in the subsequent betagalactosidase assay, except in the case of the weakest promoter clones, where 2 ml of culture was used (after 20 fold concentration by centrifugation). These results are shown in the figure. Apparently, there are very large differences in the efficiency of the cloned promoter fragments, and together these clones cover a range of promoter activities from 0.3 units of betagalactosidase activity to more than 2000 units. More interesting is it, however, that the range is covered by small changes in activity. Therefore, these promoter fragments will allow us not only to obtain a wide range of expression of genes in *L. lactis*, but also to tune the expression of genes in *L. lactis* in small steps for the purpose of flux optimization.

DNA sequencing of the 46 clones described above revealed that 64 % of the inserted promoter fragments carried the consensus sequences that was originally specified for the oligonucleotide design (see above), whereas the sequence of the remaining fragments deviated slightly from that sequence. Most of the promoter fragments that gave low activities in the betagalactosidase assay (70 units of

betagalactosidase or less) had either (A) no promoter fragment inserted, (B) an error in one of the consensus sequences, (C) a different length of spacer between the consensus sequences. All the clones that gave high activities (more than 70 units) had the same sequence as specified by the oligonucleotide. This result is in accordance with the "dogma", i.e. that changes in the consensus sequences have strong effects on the activity of a given promoter, and emphasizes the fact that a more subtle approach is needed in order to generate a promoter library that covers a range of activities in small steps. Clearly, if we would have allowed only changes in the consensus sequences and/or changes in the length of the spacer, instead of allowing the sequence of the spacers to vary randomly, only fairly weak promoter clones would have resulted, and the resulting library would not be suitable for the present purpose.

Most often, for metabolic engineering purposes, relatively strong promoters are desired, but there may also be cases where rather weak promoters are needed. The relatively few errors that had occurred during synthesis of the above oligonucleotide mixture, were not intended to be present in the promoter fragments; but our data suggest that it may sometimes be useful to generate, deliberately, a mixture of oligonucleotides so that (statistically) each oligonucleotide have a few errors in the consensus sequences. If only weak promoters are needed, it may also be useful to use an oligonucleotide mixture in which the length of spacer is shorter than what is considered to be optimal or in which parts of the consensus sequences have been changed deliberately.

Enzymes used in the various enzymatic reactions above were obtained from and used as recommended by Pharmacia and Boehringer.

EXAMPLE 2

Design of a degenerated oligonucleotide for a library of temperature regulated *L. lactis* promoters.

This example illustrates the development of a temperature regulated promoter library for *L. lactis*. Heatshock regulated promoters in Gram positive bacteria seem to have a common DNA sequence (located a few base pairs upstream of the -35 sequence) which is thought to be involved in the observed temperature regulation of the expression from such promoters. The minimal extent of such a regulatory element seems to be 27 basepairs:

5'-TTAGCACTCNNNNNNNNNGAGTGCTAA-3'
IR spacer IR

containing a 9 bp (or longer) inverted repeat (IR) separated by 9 (or fewer) basepairs. It should therefore be possible to combine this inverted repeat with the approach for obtaining constitutive promoters of different strength and thus obtain a series of promoters with various basal activities which can be induced several fold by changing the temperature of the culture medium.

Therefore, an oligonucleotide was designed, which includes the core part (from position -35 to +6) of the sequence from the constitutive promoter series above (see example 1 and SEQ ID No. 1). The sequence upstream of the -35 hexamer has been replaced by the above inverted repeat sequence, and the sequence downstream of position +6 has also been modified, eliminating two conserved regions compared to example 1 (+1 to +8: GTACTGTT and -41 to -44: AGTT, which have been implicated in growth rate regulation). The sequence of the spacer (sp.1) between the two inverted DNA sequences in the inverted repeat was here allowed to vary randomly in order to see whether this had

any effect on the temperature regulation of the resulting promoters, e.g. how many fold they could be induced by changing the temperature. The importance of the spacing (sp.2) between the inverted repeat and the -35 hexamer is not known, but in principle this region may contribute to or modulate the heatshock response of promoters. In order to limit the number of parameters, however, we have chosen here to include a naturally occuring configuration (the *dnaJ* promoter from *L. lactis*; van Asseldonk *et al.*, 1993): a short spacer sequence consisting of 5 times T.

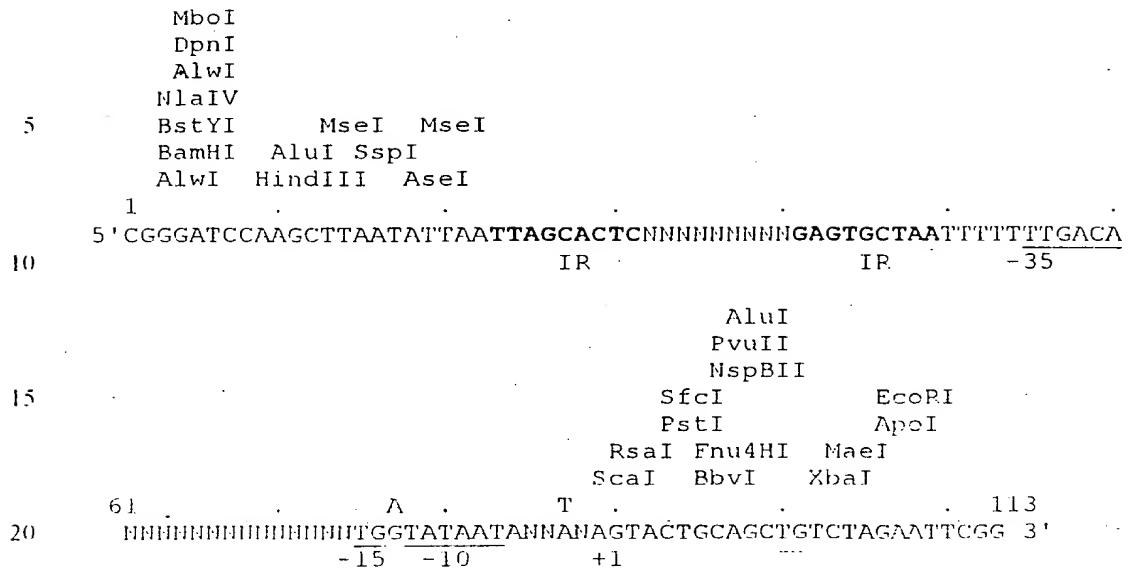
When these sequences are combined, one arrives at the following 73 bp "consensus" sequence for a temperature regulated promoter in *L. lactis*. Out of these 73 bp, 45 are conserved, two are semi-conserved (R and W) and 26 are allowed to vary randomly between the four nucleobases.

5'TTAGCACTCNNNNNNNNNGAGTGCTAATTTTTTTTGACANNNNNNNNNNNNNNNTGR

IR sp.1 IR sp.2

TATAATANNWNAGTACTG 3'

In addition, this degenerated sequence was flanked by sequences that specify multiple recognition sites for restriction endonucleases (multiple cloning sites MCS), and the actual oligonucleotide mixture that is being synthesized has the following degenerated sequence reported in SEQ ID No. 2:



This oligonucleotide mixture was converted into double stranded DNA fragments (DSDF) cloned into the promoter cloning vector pAK80 as decribed in example 1, except that the DSDF mixture and the vector pAK80 were both digested with *HindIII* and *PstI*.

EXAMPLE 3

Design of a degenerated oligonucleotide for a *Saccharomyces cerevisiae* promoter library.

The regulation of transcription initiation in the eukaryotic cell is somewhat more complex compared to the prokaryote. The transcription start site is normally preceded by a so-called TATA box that contains the consensus sequence TATAAA or parts hereof, but unlike in the prokaryote, the distance from the TATA box to the transcription start site is much less defined. In *Saccharomyces cerevisiae* this distance is typically 40-120 nucleotides (Oliver and Warmington, 1989). The so-called -35 consensus hexamer which is found in many prokaryotic promoters is absent in *Saccharomyces cerevisiae* and instead so-called upstream activation sequences (UAS) are found some

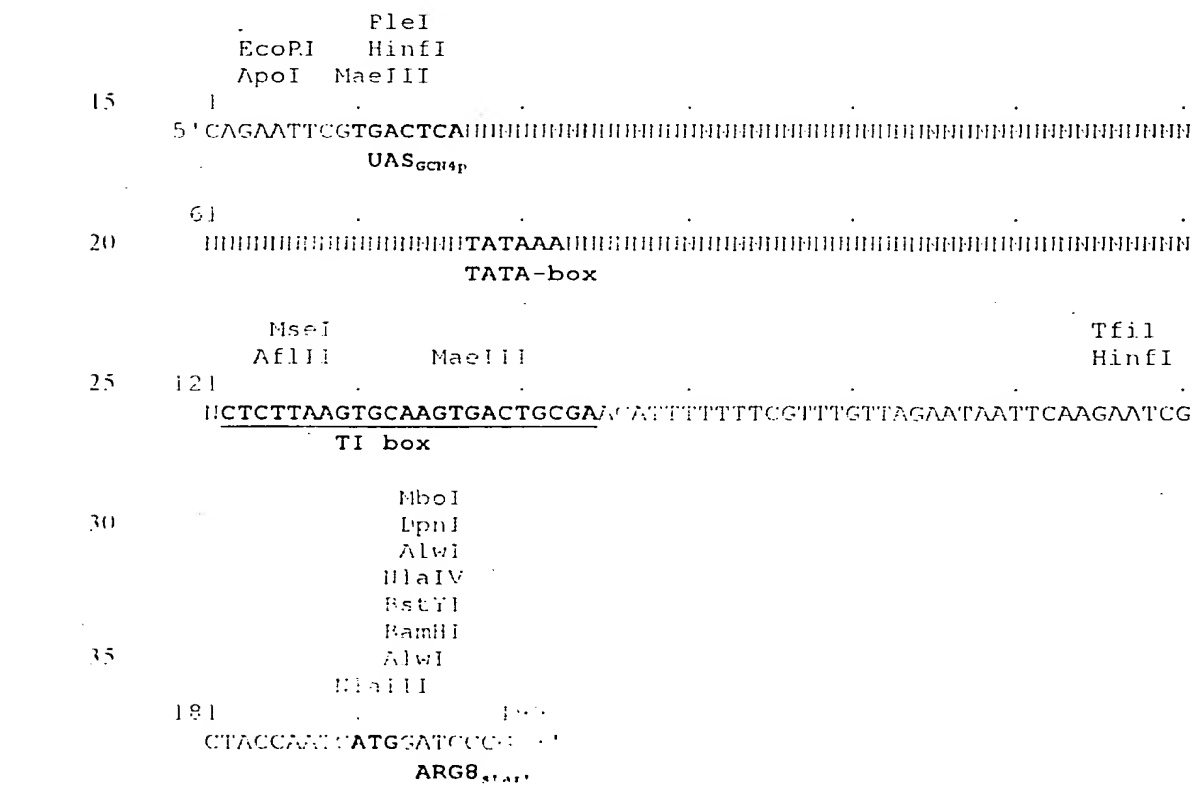
50-500 bp upstream of the transcription initiation site. These UAS are recognised by specific DNA binding proteins that can then act as activators of transcription initiation. For instance, the UAS sequence that is found upstream of the genes involved in aminoacid biosynthesis, UAS_{GCN4p}, consists of a DNA sequence that specifies a binding site for the GCN4 protein, which activates the transcription of these genes (Hinnebusch, 1986). Some genes contain more than one copy of the UAS, but one seems to be sufficient for activation. The consensus sequence for the GCN4 protein binding site is a short inverted repeat, TGACTCA.

A promoter in *Saccharomyces cerevisiae* that is activated by the GCN4 protein is the ARG8 promoter. In this promoter, there is only one copy of the UAS_{GCN4p} sequence, and it is located 59 bp from the TATA box (we refer to this distance as spacer 1). Transcription initiation takes place within a 23 bp sequence (TI box, underlined) which is located 38 bp downstream of the TATA box. This 23 bp sequence also functions as the binding site for the arginine repressor, *argR* (Crabeel *et al.*, 1995). Thus, in this case the promoter is located within 136 bp, which makes the system attractive for developing a promoter library with the degenerated oligo nucleotide approach described in example 1 and 2.

A 199 bp oligonucleotide was designed, which includes, starting from the 5' end: an *EcoRI* restriction site (for use in the subsequent cloning strategy, see below), the consensus UAS_{GCN4p}, a 60 bp spacer (spacer 1), the consensus TATAAA sequence (TATA box), a 39 bp spacer (spacer 2), the 23 bp repressor binding sequence and TI box. The spacer sequence between the TI box and the first codon of the ARG8 gene, might also contribute to the strength of the resulting promoters. However, in order to limit the number of parameters, we have chosen here to include the

native sequence from the ARG8 gene for this spacer, including also the first codon of the ARG8 gene, ATG. Finally, a *BamHI* site was included, which is designed to give an in frame fusion with the betagalactosidase reporter gene, see below.

The sequences of spacer 1 and spacer 2 were allowed to vary randomly, and the actual oligonucleotide mixture that is being synthesized has the following degenerated sequence reported in SEQ ID No. 3:



This oligonucleotide mixture was converted into double stranded DNA fragments (DSDF), using an oligonucleotide complementary to the last 23 bp of the 3' end of the 199 bp degenerated oligonucleotide as described in example 1. Subsequently, it was cloned into either of the two promoter cloning vectors, pYLZ-2 and pYLZ-6 (Hermann *et al.*, 1992), as follows: the DSDF mixture and the vector were both double-digested with *EcoRI* and *BamHI*, and the DSDF

were ligated to the vector DNA. The ligation mixture was transformed into *E. coli* as described in example 1, with selection for ampicillin resistance, and a large number of clones were obtained and pooled into a library of promoter clones. Plasmid DNA was isolated from this pool and transformed into *S. cerevisiae*, with selection for the URA3 marker carried by the vector. In addition to the *ura3* mutation, the recipient yeast strain used has two mutations: one that gives constitutive expression of GCN4 and one that inactivates the repressor, *argR*. (In this strain background the promoters resulting from the 199 bp oligonucleotide should be constitutive, which facilitates the initial screening for promoters of different strength, but if the plasmids with promoters of different strength are transformed into a wildtype strain of *S. cerevisiae*, the promoters should be regulated by the GCN4 regulon and by arginine.)

The clones were screened for betagalactosidase activity as described (Guarente, 1983).

Some yeast transformants were selected for further studies, and plasmids were rescued and characterized for the promoter structures.

REFERENCES

Crabeel, M., de Rijcke, M., Seneca, S., Heimberg, H., Pfeiffer, T., and Matisova, A., 1992. Further definition of the sequence and position requirements of the arginine control element that mediates repression and induction by arginine in Saccharomyces cerevisiae. *Yeast* 11: 1367-1380

de Vos, W.M., and Simons, G., 1994. Gene cloning and expression systems in lactococci, p.52-105. *In* M.J. Gasson and W.M. de Vos (eds.), Genetics and biotechnology of

lactic acid bacteria. Blackie Academic & Professional, Glasgow, United Kingdom.

Gasson, M.J. 1983. Plasmid complements of Streptococcus lactis NCDO 712 and other lactic streptococci after protoplast-induced curing. *J. Bacteriol.*, 154, 1-9.

Guarente, L., 1983. Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. *Methods Enzymol.* 101, 181-191.

Hermann, H., Hacker, U., Bandlow, W., and Magdolen, V., 1992. pYLZ vectors: Saccharomyces cerevisiae/Escherichia coli shuttle plasmids to analyze yeast promoters. *Gene* 119: 137-41.

Hinnebusch, A.G., 1992. General and Pathway-specific Regulatory Mechanisms Controlling the Synthesis of Amino Acid Biosynthetic Enzymes in *Saccharomyces cerevisiae*, p. 319-414. *In* E.W. Jones, J.R. Pringle and J.R. Broach (eds.), The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Gene Expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Holo, H., and Nes, I.F., 1989. High frequency transformation, by electroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. *Appl. Environ. Microbiol.*, 55, 3119-3123.

Israelsen, H. 1995. Cloning and partial characterization of regulated promoters from Lactococcus lactis Tn917-lacZ integrants with the new promoter probe vector, pAK80. *Appl. Environ. Microbiol.*, 61, 2540-2547.

Jensen, P.R., Westerhoff, H.V., and Michelsen, O., 1993. Excess capacity of H^+-ATPase and inverse respiratory control in *Escherichia coli*. *EMBO J.*, 12, 1277-1282.

Kacser, H. and Burns, J.A. 1973. The control of flux. *Symp: Soc. Exp. Biol.*, 27: 65-104.

Maniatis, T., Fritsch, E.F., and Sambrook, J., 1982. *Molecular cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Miller, J.H., 1972. *Experiments in molecular genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Nilsson, D., and Johansen, E., 1994. A conserved sequence in tRNA and rRNA promoters of Lactococcus lactis. *Biochim. Biophys. Acta*, 1219: 141-144.

Oliver, S.G., and Warmington, J.R., 1989. Transcription. In *The yeasts, volume 3*, Rose and Harrison (Eds.). Academic Press, London, 117-160.

Schaaff, I., Heinisch, J., Zimmermann, F.K., 1989. Overproduction of glycolytic enzymes in yeast. *Yeast*, 5: 285-290.

van Asseldonk, M., Simons, A., Visser, H., de Vos, W.M., and Simons, G., 1993. Cloning, nucleotide sequence, and regulatory analysis of the Lactococcus lactis dnaJ gene. *J. Bacteriol.* 175, 1637-1644.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Peter Ruhdal Jensen
(B) STREET: Soegaardsvej 19
(C) CITY: Gentofte
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): DK-2820

(ii) TITLE OF INVENTION: Artificial promoter libraries for microorganisms

(iii) NUMBER OF SEQUENCES: 3

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Lactococcus lactis

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION:26..82
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/evidence= EXPERIMENTAL
/standard_name= "Artificial promoter library"
/note= "A degenerated sequence specifying a mixture of artificial promoters covering a wide range of expression in small steps in L. lactis"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:31..45
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:60..69
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:74..82
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: -35_signal
(B) LOCATION:40..45
(D) OTHER INFORMATION:/standard_name= "-35 box"

(ix) FEATURE:
(A) NAME/KEY: -10_signal
(B) LOCATION:63..68
(D) OTHER INFORMATION:/standard_name= "Pribnow box"

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:3..25
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/evidence= EXPERIMENTAL
/standard_name= "Multiple cloning site"
/label= MCS
/note= "A sequence specifying recognition sites for the restriction endonucleases: NlaIV, BstYI, BamHI, AlwI, MboI,DpnI, AflII, MseI, SspI, NsiI."

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:74..98
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/evidence= EXPERIMENTAL
/standard_name= "Multiple cloning site"
/label= MCS
/note= "A sequence specifying recognition sites for the restriction endonucleases: ScaI, RsaI, HpaI, HincII, MseI, SfcI, PstI, Fnu4HI, BbvI, PvuII, NspBII, AluI, EcoRI."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CGGGATCCTT AAGAATATTA TGCATNNNNN AGTTTATTCT TGACANNNNN NNNNNNNNNT 60

GGTATAATAN NANAGTACTG TTAACTGCAG CTGAATTCGG 100

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION:23..95
(D) OTHER INFORMATION:/standard_name= "Artificial promoter library"
/note= "A degenerated sequence specifying a mixture of artificial temperature regulated promoters covering a wide range of expression in small steps in L. lactis"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:23..49
(D) OTHER INFORMATION:/standard_name= "Sequence providing temperature regulation to promoters"
/note= "This sequence comprising two inverted repeats separated by a short spacer provides temperature (heat shock) regulation to promoters in Gram-positive bacteria"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:50..60
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:75..84
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:89..95
(D) OTHER INFORMATION:/standard_name= "Consensus sequence"

(ix) FEATURE:
(A) NAME/KEY: -35_signal
(B) LOCATION:55..60
(D) OTHER INFORMATION:/standard_name= "-35 box"

(ix) FEATURE:
(A) NAME/KEY: -10_signal
(B) LOCATION:78..83
(D) OTHER INFORMATION:/standard_name= "Pribnow box"

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:3..22
(D) OTHER INFORMATION:/standard_name= "Multiple cloning site"
/label= MCS
/note= "A sequence specifying recognition sites for the restriction endonucleases: NlaIV, BstYI, BamHI, AlwI, MboI, DpnI, HindIII, AluI, MseI (2 sites), SspI, AseI"

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:89..111
(D) OTHER INFORMATION:/standard_name= "Multiple cloning site"
/label= MCS
/note= "A sequence specifying recognition sites for the restriction endonucleases: ScaI, RsaI, SfcI, PstI, Fnu4HI, BbvI,PvuII, NspBII, AluI, XbaI, MaeI, EcoRI, ApoI"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CGGGATCCAA GCTTAATATT AATTAGCACT CNNNNNNNNN GAGTGCTAAT TTTTTTGACA 60

NNNNNNNNNN NNNNTGGTAT AATANNANAG TACTGCAGCT GTCTAGAATT CGG 113

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: protein_bind
(B) LOCATION:10..16
(D) OTHER INFORMATION:/function= "Activating promoters in S. cerevisiae"
/bound_moiety= "GCN4 protein"
/standard_name= "Upstream activating sequence"
/label= UAS_GCN4p
/note= "A DNA sequence that specifies a binding site for the GCN4 protein, which activates the transcription of genes involved in aminoacid synthesis in S. cerevisiae"

(ix) FEATURE:
(A) NAME/KEY: TATA_signal
(B) LOCATION:67..72
(D) OTHER INFORMATION:/standard_name= "TATA box"

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION:122..144
(D) OTHER INFORMATION:/function= "Transcription initiation"
/standard_name= "TI box"

(ix) FEATURE:
(A) NAME/KEY: protein_bind
(B) LOCATION:122..144
(D) OTHER INFORMATION:/bound_moiety= "Arginine repressor"
/standard_name= "arginine repressor binding site"
/label= argR

(ix) FEATURE:
(A) NAME/KEY: misc_RNA
(B) LOCATION:1[illegible]..192
(D) OTHER INFORMATION:/function= "Spacer"
/standard_name= "Part of native sequence for ARG8 gene in [illegible] first codon"

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:[illegible]
(D) OTHER INFORMATION:/standard_name= "Recognition site for restriction endonuclease EcoRI"
/label= EcoRI_site

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION:19[illegible]..197
(D) OTHER INFORMATION:/standard_name= "Recognition site for restriction endonuclease BamHI"
/label= BamHI_site

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION:10..192
(D) OTHER INFORMATION:/standard_name= "Artificial promoter library"
/note= "A degenerated sequence specifying a mixture of artificial promoters covering a wide range of expression in small steps in S. cerevisiae"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CAGAATTCGT GACTCACTTA AAGTACATGA TCCGTCATTG TGCACTTTTT TACGAAATAA 60

TGGCCTTTTT CCGCTCTATT TAAAAGCGTG AAAAAAAAAT TGCAACATGA AGAAAATTCG 120

ACTCTTAAGT GCAAGTGACT GCGAACATTT TTTTCGTTTG TTAGAATAAT TCAAGAATCG 180

CTACCAATCA TGGATCCCG 199

PATENT CLAIMS

1. An artificial promoter library for a selected microorganism or group of microorganisms, comprising a mixture of double stranded DNA fragments the sense strands of which comprise at least one consensus sequence or part(s) thereof of efficient promoters from said microorganism or group of microorganisms and surrounding or intermediate nucleotide sequences (spacers) of variable length in which the nucleotides are selected randomly among the nucleobases A, T, C and G.

2. An artificial promoter library according to claim 1, wherein the sense strands of the double stranded DNA fragments include a regulatory DNA sequence imparting a specific regulatory feature to the promoters of the library.

3. An artificial promoter library according to claim 2, wherein the specific regulatory feature is activation by a change in the growth conditions.

4. An artificial promoter library according to any one of claims 1-3, wherein the double stranded DNA fragments have sequences comprising one or more recognition sites for restriction endonucleases added to their ends.

5. An artificial promoter library according to claim 4, wherein the double stranded DNA fragments have sequences specifying multiple recognition sites for restriction endonucleases (multiple cloning sites MCS) added to their ends.

6. An artificial promoter library according to any one of claims 1-5, wherein the selected microrganism or group

of microorganisms is selected from the group consisting of prokaryotes.

7. An artificial promoter library according to claim 6, wherein the selected microrganism or group of microorganisms is selected from the group consisting of lactic acid bacteria of the genus *Lactococcus*.

8. An artificial promoter library according to claim 7, wherein the selected microorganism is a strain of the species *Lactococcus lactis*.

9. An artificial promoter library according to claim 7 or 8, wherein said consensus sequences comprise the -35 signal (-35 to -30): TTGACA or the -10 signal (-12 to -7): TATAAT or a part of one of them comprising at least 4 conserved nucleotides.

10. An artificial promoter library according to claim 7 or 8, wherein said consensus sequences comprise the -35 signal (-35 to -30): TTGACA and the -10 signal (-12 to -7): TATAAT or parts of both comprising together at least 6 conserved nucleotides.

11. An artificial promoter library according to claim 9 or 10, wherein said consensus sequences further comprise intervening conserved motifs.

12. An artificial promoter library according to claim 11, wherein said intervening conserved motifs are selected from the group consisting of the conserved motifs -44 to -41: AGTT, -40 to -36: TATTC and +1 to +8: GTACTGTT.

13. An artificial promoter library according to any one of claims 10-12, wherein the length of the spacer between the -35 signal and the -10 signal is 14-23 bp.

14. An artificial promoter library according to claim 13, wherein the length of the spacer between the -35 signal and the -10 signal is 16-18 bp.

15. An artificial promoter library according to claim 13, wherein the length of the spacer between the -35 signal and the -10 signal is 17 bp.

16. An artificial promoter library according to any one of claims 12-15, wherein the sense strands of the double stranded DNA fragments have the degenerated sequence shown in SEQ ID No. 1 with minor variations in the consensus sequences and spacer lengths.

17. An artificial promoter library according to claim 10, wherein the sense strands of the double stranded DNA fragments, including an upstream temperature regulatory DNA sequence, have the degenerated sequence shown in SEQ ID No. 2 with minor variations in the consensus sequences and spacer lengths.

18. An artificial promoter library according to any one of claims 1-5, wherein the selected microrganism or group of microorganisms is selected from the group consisting of eukaryotic microorganisms.

19. An artificial promoter library according to claim 18, wherein the selected microrganism or group of microorganisms is selected from the group consisting of yeasts, other fungi and mammalian cell lines.

20. An artificial promoter library according to claim 19, wherein the selected microorganism is a strain of the yeast species *Saccharomyces cerevisiae*.

21. An artificial promoter library according to claim 20, wherein said consensus sequences comprise a tran-

scription initiation signal (TI box) functioning in *Saccharomyces cerevisiae*.

22. An artificial promoter library according to claim 21, wherein said consensus sequences further comprise a TATA box and/or at least one upstream activation sequence (UAS) upstream of the TI box.

23. An artificial promoter library according to claim 22, wherein said consensus sequences comprise the TI box: CTCTTAAGTGCAAGTGACTGCGA, which also functions as the binding site for the arginine repressor, *argR*, the TATA box: TATAAA, and the UAS_{GCN4p}: TGACTCA.

24. An artificial promoter library according to claim 23, wherein the sense strands of the double stranded DNA fragments have the degenerated sequence shown in SEQ ID No. 3 with minor variations in the consensus sequences and spacer lengths.

25. A method of constructing an artificial promoter library for a selected microorganism or group of microorganisms, which comprises selecting at least one consensus sequence of efficient promoters from said microorganism or group of microorganisms; synthesizing a mixture of single stranded DNA sequences comprising said consensus sequence(s) or part(s) thereof and surrounding or intermediate nucleotide sequences (spacers)of variable length in which the nucleotides are selected randomly among the nucleobases A, T, C and G; and converting the single stranded DNA sequences into double stranded DNA fragments by second strand synthesis.

26. A method according to claim 25, wherein the single stranded DNA sequences include a regulatory DNA sequence imparting a specific regulatory feature to the promoters of the library.

27. A method according to claim 26, wherein the specific regulatory feature is activation by a change in the growth conditions

28. A method according to any one of claims 25-27, wherein sequences specifying one or more recognition sites for restriction endonucleases are added to the ends of the single stranded DNA sequences in the synthesis, or linkers comprising such restriction sites are ligated to the ends of the double stranded DNA fragments.

29. A method according to claim 28, wherein sequences specifying multiple recognition sites for restriction endonucleases are added to the ends of the single stranded DNA sequences in the synthesis, or linkers comprising such multiple cloning sites (MCS) are ligated to the ends of the double stranded DNA fragments.

30. A method according to any one of claims 25-29, wherein the selected microrganism or group of microorganisms is selected from the group consisting of prokaryotes.

31. A method according to claim 30, wherein the selected microrganism or group of microorganisms is selected from the group consisting of lactic acid bacteria of the genus *Lactococcus*.

32. A method according to claim 31, wherein the selected microorganism is a strain of the species *Lactococcus lactis*.

33. A method according to claim 31 or 32, wherein said consensus sequences comprise the -35 signal (-35 to -30): TTGACA or the -10 signal (-12 to -7): TATAAT or a part of one of them comprising at least 4 conserved nucleotides.

34. A method according to claim 31 or 32, wherein said consensus sequences comprise the -35 signal (-35 to -30): TTGACA and the -10 signal (-12 to -7): TATAAT or parts of both comprising together at least 6 conserved nucleotides.

35. A method according to claim 33 or 34, wherein said consensus sequences further comprise intervening conserved motifs.

36. A method according to claim 35, wherein said intervening conserved motifs are selected from the group consisting of the conserved motifs -44 to -41: AGTT, -40 to -36: TATTC and +1 to +8: GTACTGTT.

37. A method according to any one of claims 34-36, wherein the length of the spacer between the -35 signal and the -10 signal is 14-23 bp.

38. A method according to claim 37, wherein the length of the spacer between the -35 signal and the -10 signal is 16-18 bp.

39. A method according to claim 37, wherein the length of the spacer between the -35 signal and the -10 signal is 17 bp.

40. A method according to any one of claims 34-39, wherein the mixture of single stranded DNA sequences has the degenerated sequence shown in SEQ ID No. 1 with minor variations in the consensus sequences and spacer lengths.

41. A method according to any one of claims 34, 35 and 37-39, wherein the mixture of single stranded DNA sequences including an upstream temperature regulatory DNA sequence has the degenerated sequence shown in SEQ ID No.

2 with minor variations in the consensus sequences and spacer lengths.

42. A method according to any one of claims 25-29, wherein the selected microrganism or group of microorganisms is selected from the group consisting of eukaryotic microorganisms.

43. A method according to claim 42, wherein the selected microrganism or group of microorganisms is selected from the group consisting of yeasts, other fungi and mammalian cell lines.

44. A method according to claim 43, wherein the selected microorganism is a strain of the yeast species *Saccharomyces cerevisiae*.

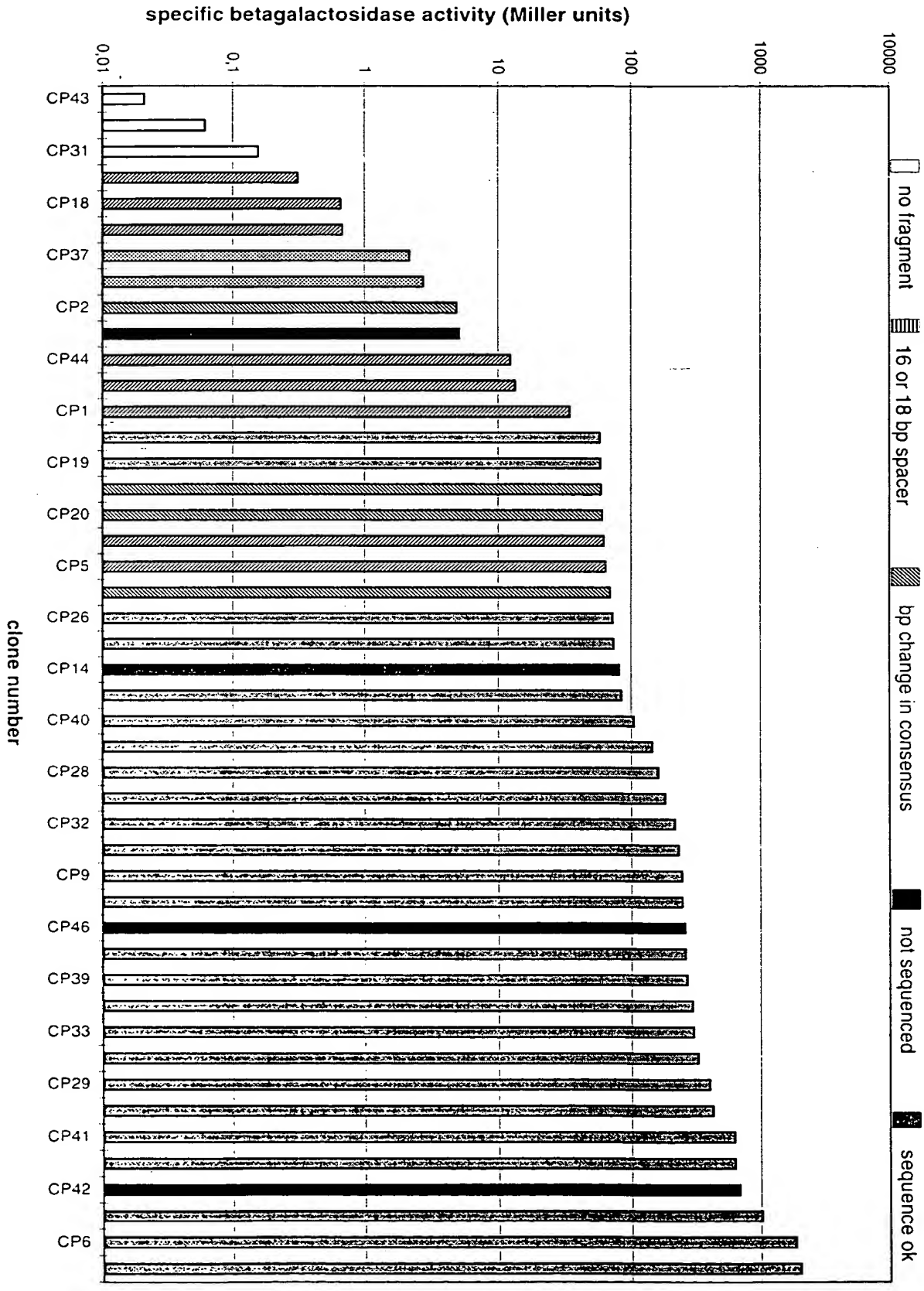
45. A method according to claim 44, wherein said consensus sequences comprise a transcription initiation signal (TI box) functioning in *Saccharomyces cerevisiae*.

46. A method according to claim 45, wherein said consensus sequences further comprise a TATA box and/or at least one upstream activation sequence (UAS) upstream of the TI box.

47. An artificial promoter library according to claim 46, wherein said consensus sequences comprise the TI box: CTCTTAAGTGCAAGTGACTGCGA, which also functions as the binding site for the arginine repressor, *argR*, the TATA box: TATAAA, and the UAS_{GCN4p}: TGACTCA.

48. An artificial promoter library according to claim 47, wherein the sense strands of the double stranded DNA fragments have the degenerated sequence shown in SEQ ID No. 3 with minor variations in the consensus sequences and spacer lengths.

49. A method of optimizing the expression of a recombinant heterologous gene in a selected microorganism, which comprises transforming or transfecting the microorganism with a set of vectors each including said heterologous gene under the control of at least one member of an artificial promoter library according to any one of claims 1-24 or constructed by the method according to any one of claims 25-48, said set of vectors covering a wide range of promoter activities in relatively small steps, growing the selected clones and screening them to find the one showing maximum metabolic flux of the product encoded by said heterologous gene.


specific betagalactosidase activity (Miller units)
0,01
0,1
1
10
100
1000
10000
clone number
CP43
CP31
CP18
CP37
CP2
CP44
CP1
CP19
CP20
CP5
CP26
CP14
CP40
CP28
CP32
CP9
CP46
CP39
CP33
CP29
CP41
CP42
CP6
no fragment
16 or 18 bp spacer
bp change in consensus
not sequenced
sequence ok

Hofman-Bang & Boutard, Lehmann & Ree A/S
Patents Trademarks Designs

Patentdirektoratet
Rettighedsafdelingen
Patenter
Helgeshøj Allé
2630 Tåstrup

Deres ref.:	Vor ref.:	Dato:
-	2960322 JNY/LLN	22. oktober 1996

Ans. nr.: 0886/96 - P2
Direktoratets brev af 06 september 1996
Frist: 06 december 1996

Gebyr: Kr. -

I den ovennævnte ansøgning, som blev indleveret den 23. august 1996, er vi blevet opmærksomme på nogle åbenbare fejl, som vi hermed ønsker at rette:

I SEQUENCE LISTING, (2) INFORMATION FOR SER ID NO:3, (xi) SEQUENCE DESCRIPTION: SEQ ID NO.3 på side 27 var der ved en fejltagelse indført et udgangsoligonucleotid før degenereringen af sekvensen som angivet i Example 3. Da den degenererede sekvens, som skal være SEQ ID No 3, tydeligt er angivet i Example 3 på side 19, har vi nu indført denne korrekte SEQ ID No 3 i SEQUENCE LISTING på side 27.

Desuden var krav 47 og 48, som var kopieret fra krav 23 og 24, ved en fejltagelse ikke blevet rettet på "A method" At de skulle være det, fremgår imidlertid klart af, at de henviser til hhv. krav 46 og 47, og vi har derfor nu rettet "An artificial promoter library" til "A method" i krav 47 og 48.

Vi vedlægger en ny side 27 med den rigtige SEQ ID No. 3 og en ny side 34, hvor krav 47 og 48 er rettet som forklaret ovenfor

Desuden vedlægges en diskette med sekvenslisten efter Patentin-systemet som tekstfil: 2960322.seq.

☑ Adelgade 15
DK-1304 København K
Danmark
Telefon: +45 33 15 05 85
Telefax: +45 33 15 75 85
Telex: 19 085 hbb dk

☐ Grundtvigsvej 37
DK-1864 Frederiksberg C
Danmark
Telefon: +45 31 31 32 31
Telefax: +45 31 31 12 90
Telex: 15 067 lehree dk

☐ Ryesgade 3
Postboks 367
DK-8100 Aarhus C
Danmark
Telefon: +45 86 20 22 22
Telefax: +45 86 20 22 10

Telegram: Hofmanbang
Reg. nr.: 207 235
Giro: 7 02 57 85

Endelig vedlægges den manglende overdragelseserklæring.

Med venlig hilsen
Hofman-Bang & Boutard, Lehmann & Ree A/S

Jørgen Nyeng

Bilag
Ny side 27 in duplo
Ny side 34 in duplo
Diskette
Overdragelseserklæring

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CAGAATTCGT GACTCANNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN 60

NNNNNNNNNN NNNNNNTATA AANNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN 120

NCTCTTAAGT GCAAGTGACT GCGAACATTT TTTTCGTTTG TTAGAATAAT TCAAGAATCG 180

CTACCAATCA TGGATCCCG 199

2 with minor variations in the consensus sequences and spacer lengths.

42. A method according to any one of claims 25-29, wherein the selected microrganism or group of microorganisms is selected from the group consisting of eukaryotic microorganisms.

43. A method according to claim 42, wherein the selected microrganism or group of microorganisms is selected from the group consisting of yeasts, other fungi and mammalian cell lines.

44. A method according to claim 43, wherein the selected microorganism is a strain of the yeast species *Saccharomyces cerevisiae*.

45. A method according to claim 44, wherein said consensus sequences comprise a transcription initiation signal (TI box) functioning in *Saccharomyces cerevisiae*.

46. A method according to claim 45, wherein said consensus sequences further comprise a TATA box and/or at least one upstream activation sequence (UAS) upstream of the TI box.

47. A method according to claim 46, wherein said consensus sequences comprise the TI box:
CTCTTAAGTGCAAGTGACTGCGA, which also functions as the binding site for the arginine repressor, *argR*, the TATA box: TATAAA, and the UAS_{GCN4p}: TGACTCA.

48. A method according to claim 47, wherein the sense strands of the double stranded DNA fragments have the degenerated sequence shown in SEQ ID No. 3 with minor variations in the consensus sequences and spacer lengths.